Supplementary Information

A divide–and–conquer approach
to analyze underdetermined biochemical models

Oliver Kotte & Matthias Heinemann

Contents

Verbose derivation of Conditions 1–3

When fulfilled, Conditions 1–3, i.e. Equations (8), (10) and (12) of the main paper, decompose the global parameter estimation problem into smaller subproblems. The derivation of these conditions is fairly condensed in the main paper due to space limitations. Therefore, we here provide a more verbose derivation.

In Section 2 of the main paper, we depart from a general formulation of the global estimation problem, presented in Equations (1)–(5), and successively specialize this general formulation to a formulation composed of independent subproblems, Eq. (13). The conditions imposed during this specialization are Conditions 1–3, the necessary and sufficient conditions to trigger the decomposition.

The first thing we do in the main paper is to rewrite Eq. (2) to Eq. (6) because the latter equation explicitly contains the observable rates v. Then, to get rid of the integral in Eq. (1), we take advantage of the discrete nature of biochemical measurements and replace the integral with a sum over the measurement time points, resulting in Eq. (7). The general formulation of the global estimation problem that we next specialize is thus given by Equations (3)–(7).

The first condition we impose to specialize the general formulation is to demand that at all measurement time points t_i, the measurement data set must consist of all differential state variables x and all rates v. Therefore, $\mathbf{y}_{msd}^T = \left(\mathbf{x}_{msd}^T \, \mathbf{v}_{msd}^T\right)$ (Eq. (8)). To be able to calculate the difference between measurement and model prediction, $\mathbf{y}_{msd}(t_i) - \mathbf{y}(t_i)$,

which appears in the cost function (Eq. (7)), the model must also predict all differential state variables x and all rates v at all measurement time points t_i. Therefore, the predictor function must be $\mathbf{k}^T = \left(\mathbf{x}^T\ \mathbf{v}^T\right)$ (Eq. (9)). Because of Eq. 3, this predictor function leads to the required model prediction $\mathbf{y} = \left(\mathbf{x}^T\ \mathbf{v}^T\right)$.

Next, we plug Equations (8) and (9) into the cost function, Eq. (7). Therefore, in Eq. (7), $\mathbf{y}_{msd}$ becomes $\left(\mathbf{x}_{msd}^T\ \mathbf{v}_{msd}^T\right)^T$ and $\mathbf{y}$ becomes $\left(\mathbf{x}^T\ \mathbf{v}^T\right)^T$. This formulation of the cost function J appears in Eq. (10). We next make the argument that $J = 0$ must be a global optimum because the cost function J is by definition non–negative. $J \geq 0$ must hold because J is a sum–of–squares, which is obviously non–negative, weighted by the scaling matrix $\mathbf{W}$, which cannot introduce a sign change due to being diagonal with non–negative elements. For $J = 0$ to occur, the model structure must allow for an exact fit to the data, which in turn essentially requires an underdetermined problem. The second condition we impose is to demand that $J = 0$ does indeed occur, as stated in Eq. (10).

We next exploit the fact that a sum–of–squares can only vanish when all the individual summands vanish. Because of this, Eq. (10) can only hold if $\mathbf{x}_{msd}(t_i) = \mathbf{x}(t_i)$ and $\mathbf{v}_{msd}(t_i) = \mathbf{v}(t_i)$ for all $t_i\ , i = 1 \ldots m$. From Eq. (6), we also know that $\mathbf{v}$ is a function of $(\mathbf{x}, \mathbf{p}, \mathbf{q}, t)$. Therefore, we know that $\mathbf{v}\left(\mathbf{x}, \mathbf{p}, \mathbf{q}, t_i\right) = \mathbf{v}(t_i)_{msd}$. Because of $\mathbf{x}_{msd}(t_i) = \mathbf{x}(t_i)$, as stated above, we arrive at $\mathbf{v}\left(\mathbf{x}_{msd}(t_i), \mathbf{p}, \mathbf{q}, t_i\right) = \mathbf{v}_{msd}(t_i)$, which is the vector formulation of Eq. (11).

With r the number of rates or components in $\mathbf{v}$, Eq. (11) comprises $m{\cdot}r$ equations. The only unknown in Eq. (11) is $\mathbf{p}$. Therefore, Eq. (11) is a system of $m{\cdot}r$ algebraic equations, which is in practice underdetermined and can be solved to obtain the complete solution space of $\mathbf{p}$. However, depending on the model size, the derived solution space can be very large–dimensional, and therefore nontransparent to the modeler. To understand the solution space better, it would be of tremendous help if the solution space consisted of pairwise independent, smaller–dimensional subspaces. To trigger such a decomposition of the solution space, we next exploit the fact that parameters in biochemical models tend to have a specific mechanistic meaning and as such appear only in one rate equation. We thus impose the third condition (Eq. (12)), which demands that the parameter set consists of disjunct subsets such that each of these subsets fully parameterizes a subset of the rate equations. The desired extreme of this condition is that each parameter indeed appears in only one rate equation, which leads to a maximal decomposition of the solution space. The undesired extreme is that such disjunct sets do not exist at all, which does not allow for a decomposition of the solution space.

Because of the decomposition of the parameter vector $\mathbf{p}$, which is the sole unknown in the equation system of Eq. (11), this equation system is also decomposed into s pairwise independent subsets (Eq. (13)). Therefore, Eq. (13) is a formulation of the global estimation problem that consists of s independent subproblems in the form of algebraic equation systems, which can be solved independently to obtain the complete solution spaces of the parameter vectors $\mathbf{p}_k$ associated with the subproblems. The aggregate of these solution subspaces is the complete space of global solutions to the parameter estimation problem of Equations (3)–(7), given that the three imposed conditions hold.

Model equations

This section presents the model equations of the example system introduced in the main text (see Figure 2a).

The system comprises the six dynamic state variables

$$\mathbf{x} = [E_1 \;\; E_3 \;\; E_4 \;\; E_5 \;\; PEP \;\; FBP]^T \;, \tag{S1}$$

which model the four enzyme and the two metabolite concentrations.

These compounds are interconverted by the 15 rates

$$\begin{aligned}\mathbf{v} = [&v_{ex,e_1} \;\; v_{ex,e_3} \;\; v_{ex,e_4} \;\; v_{ex,e_5} \;\; v_{d,E_1} \;\; v_{d,E_3} \;\; v_{d,E_4} \;\; v_{d,E_5} \\ &v_{d,PEP} \;\; v_{d,FBP} \;\; v_{r,E_1} \;\; v_{r,E_2} \;\; v_{r,E_3} \;\; v_{r,E_4} \;\; v_{r,E_5}]^T \;,\end{aligned} \tag{S2}$$

where the subscripts ex denote expression rates, d denote combined dilution and degradation rates, r denote metabolic reaction rates, and e_i denote the genes encoding for the modeled enzymes.

Note that we have simplified the binding process of FBP and Cra such that the binding state of Cra directly tracks the concentration of its effector metabolite FBP. Therefore,

$$Cra_A = \frac{Cra_T}{1 + \left(\frac{FBP}{K_{Cra,FBP}}\right)^{n_{Cra}}} \;, \tag{S3}$$

where Cra_A is the active concentration of Cra that is not bound to FBP, Cra_T is the total Cra concentration and $K_{Cra,FBP}$ the FBP concentration required for half–saturation of Cra with FBP.

With these definitions, the time progression of these state variables is

$$\begin{aligned}\tfrac{dE_i}{dt} &= v_{ex,e_i} - v_{d,E_i} \\ \tfrac{dPEP}{dt} &= v_{r,E_1} + v_{r,E_2} - v_{r,E_4} - v_{r,E_5} - v_{d,PEP} \\ \tfrac{dFBP}{dt} &= -0.5 \cdot v_{r,E_2} + v_{r,E_5} - v_{r,E_3} - v_{d,FBP} \;.\end{aligned} \tag{S4}$$

In the remainder of this section, we present the algebraic equations for all the rates v_i appearing in Eq. (S4).

If the transcription factor Cra acts as activator on the expression of an enzyme E from its gene e, the expression rate is given

$$v_{ex,e} = \rho \; \mu \; v_{e,max} \; \frac{Cra_A}{Cra_A + K_{e,Cra_A}} \;, \tag{S5}$$

and if Cra acts as repressor, the expression rate is given by

$$v_{ex,e} = \rho \; \mu \; v_{e,max} \left(1 - \frac{Cra_A}{Cra_A + K_{e,Cra_A}}\right) \;. \tag{S6}$$

In these equations, μ is the growth rate, $v_{e,max}$ the maximal expression rate, and K_{e,Cra_A} the active Cra concentration required for half–maximal expression. ρ parameterizes the

linear influence that μ exerts on the efficiency of the gene expression machinery. ρ is set to 1, because its value is anyhow later corrected with the optimization of the multiplicative parameters $v_{e,max}$.

The combined dilution and degradation rates of a compound x are given by

$$v_{d,x} = (\mu + k_{degr})\ x, \tag{S7}$$

with a degradation rate of $k_{degr} = 0$ in the case of metabolites, and $k_{degr} = 2.8 \cdot 10^{-5} s^{-1}$ in the case of proteins (Endy *et al.*, 1996).

The enzymes E_3 and E_4 are activated in a cooperative manner by PEP and FBP, respectively, and are modeled using a Monod-Wyman-Changeux (MWC) kinetics, such that

$$v_{r,E_i} = \frac{k_{cat,E_i}\ E_i\ \frac{S}{K_{E_i,S}} \left(1 + \frac{S}{K_{E_i,S}}\right)^{n_{E_i}-1}}{\left(\frac{S}{K_{E_i,S}}\right)^{n_{E_i}} + L_{E_i}\ /\ \left(1 + \frac{M}{K_{E_i,M}}\right)^{n_{E_i}}}\ . \tag{S8}$$

The enzymes E_1 and E_5 are inhibited in a cooperative manner by FBP and PEP, respectively, and are also modeled with a MWC kinetics, such that

$$v_{r,E_i} = \frac{k_{cat,E_i}\ E_i\ \frac{S}{K_{E_i,S}} \left(1 + \frac{S}{K_{E_i,S}}\right)^{n_{E_i}-1}}{\left(\frac{S}{K_{E_i,S}}\right)^{n_{E_i}} + L_{E_i}\ \left(1 + \frac{M}{K_{E_i,M}}\right)^{n_{E_i}}}\ . \tag{S9}$$

In these equations, k_{cat,E_i} denote the maximal turnover capacities, $K_{E_i,S}$ and $K_{E_i,M}$ the required respective concentrations of substrate and effector for half–saturation, n_{E_i} the number of monomers in the active enzyme complex, and L_{E_i} parameterizes the ligand binding. The enzyme E_2 is modeled with a reversible Michaelis–Menten kinetics, such that

$$v_{r,E_2} = \frac{v_{E_2,f}\ E_2\ \frac{FBP}{K_{E_2,FBP}} - v_{E_2,r}\ E_2\ \frac{PEP}{K_{E_2,PEP}}}{1 + \frac{FBP}{K_{E_2,FBP}} + \frac{PEP}{K_{E_2,PEP}}}, \tag{S10}$$

where $v_{E_2,f}$ and $v_{E_2,r}$ are the maximal reaction rates in the forward and reverse direction, respectively, and $K_{E_2,FBP}$ and $K_{E_2,PEP}$ the substrate concentrations required for half–saturation. Note that this enzyme's concentration is assumed to be constant.

Measurement data set

Table S1 lists the measurement data (Fuhrer *et al.*, 2005; Lowry *et al.*, 1971; Oh *et al.*, 2002; Zhao *et al.*, 2004) that were used to optimize the parameters. The relative enzyme concentrations listed therein were derived from DNA microarray data, thereby assuming that enzyme concentrations scale with their mRNA concentrations.

Table S1: Measurement data obtained for balanced growth on glucose and acetate.

Condition	$Acetate$	$Glucose$	μ	PEP	FBP
Acetate	5	0	0.20	0.59	0.28
Glucose	0	5	0.64	0.21	6.6
Condition	E_1	E_2	E_3	E_4	E_5
Acetate	10.65	1	3.3	0.61	0.59
Glucose	1	1	1	1	1
Condition	J_{E_1}	J_{E_2}	J_{E_3}	J_{E_4}	J_{E_5}
Acetate	0.198	-0.188	0.094	0.01	0
Glucose	0	3.871	0.06	1.874	1.997

Extracellular metabolites in $[\frac{g}{l}]$. Growth rate in $[\frac{1}{h}]$. Intracellular metabolites in $[\frac{\mu mol}{gDW}]$. Enzyme levels in [AU]. Metabolic fluxes in $[\frac{\mu mol}{gDW\ s}]$.

Derivation of the equality constraints on the parameters

This section supplements the exemplary application of the divide–and–conquer approach to obtain the complete solution space of the model defined in Equations (S1)–(S10). In the following, we derive the equality constraints on the parameters from the measurement data shown in Table S1, and show how these constraints divide the parameter set $\mathbf{p}$ into free and dependent parameters. The equality constraints are derived independently for each of the six subproblems into which the global optimization problem decomposes (the six subproblems are summarized in Table 1 and Figure 2b of the main text). In the following, the subscripts $_{Glc}$ and $_{Act}$ refer to measurement data points in the glucose and acetate data sets, respectively.

First, we optimize the parameters of subproblem 1, which contains the four gene expression rates and the binding of the transcription factor Cra to the metabolite FBP. Each of the four gene expression rates v_{ex,e_i} with $i = 1, 3, 4, 5$ contains two parameters, $v_{e_i,max}$ and K_{e_i,Cra_A}. With the two measurement data sets, these parameters can be determined exactly as a function of the two parameters describing the transcription factor–metabolite binding, $K_{Cra,FBP}$ and n_{Cra}. With Cra_A according to Eq. S1 and $v_{d,x}$ according to Eq. S7, the explicit solutions are

$$\begin{aligned} K_{e_i,Cra_A} &= (\nu\, Cra_{A,Glc} - Cra_{A,Act})(1-\nu)^{-1} && ,\ i = 1,3,4,5 \\ v_{e_i,max} &= v_{d,E_i,Glc}\,(Cra_{A,Glc} + K_{e_i,Cra_A})(\rho\,\mu_{Glc}\,Cra_{A,Glc})^{-1} && ,\ i = 1,3 \\ v_{e_i,max} &= v_{d,E_i,Glc}\,[\rho\,\mu_{Glc}\,(1 - Cra_{A,Glc}(Cra_{A,Glc} + K_{e_i,Cra_A})^{-1})]^{-1} && ,\ i = 4,5 \end{aligned} \tag{S11}$$

with

$$\begin{aligned} \nu &= Cra_{A,Act}\,\mu_{Act}\,v_{d,E_i,Glc}\,(Cra_{A,Glc}\,\mu_{Glc}\,v_{d,E_i,Act})^{-1} && ,\ i = 1,3 \\ \nu &= \mu_{Act}\,v_{d,E_i,Glc}\,(\mu_{Glc}\,v_{d,E_i,Act})^{-1} && ,\ i = 4,5 \quad . \end{aligned} \tag{S12}$$

These equations reduce the ten–dimensional parameter space to a two–dimensional solution space.

Next, we optimize the parameters of the five enzyme kinetics. The kinetic equations for both uptake enzymes E_1 (subproblem 2) and E_5 (subproblem 3) each contain four uncertain parameters, with only one constraining measurement each. We chose to express v_{E_i} with $i = 1, 5$ as a function of the remaining parameters, such that:

$$\begin{aligned} v_{E_1} &= \frac{J_{E_1,Act}\,K_{E_1,Acetate}}{Acetate\cdot E_{1,Act}} \cdot \frac{\left(1+Acetate/K_{E_1,Acetate}\right)^{n_1} + L_{E_1}\left(1+PEP_{Act}/K_{E_1,PEP}\right)}{\left(1+Acetate/K_{E_1,Acetate}\right)^{n_1-1}} \\ v_{E_5} &= \frac{J_{E_5,Glc}\,K_{E_5,Glucose}}{Glucose\cdot E_{5,Glc}} \cdot \frac{\left(1+Glucose/K_{E_5,Glucose}\right)^{n_5} + L_{E_5}\left(1+FBP_{Glc}/K_{E_5,FBP}\right)}{\left(1+Glucose/K_{E_5,Glucose}\right)^{n_5-1}} \end{aligned} \tag{S13}$$

These equations reduce the four–dimensional parameter spaces to three–dimensional solution spaces.

The kinetic equations for the enzymes E_3 (subproblem 4) and E_4 (subproblem 5) each contain four uncertain parameters, with two measurements to constrain them. With $i =$

3, 4, we chose to express v_{E_i} and the $K_{E_i,S}$-values of the respective substrates, $K_{E_3,PEP}$ and $K_{E_4,FBP}$, as a function of the remaining parameters. This gives two functions,

$$\begin{aligned} f_1 &= \frac{J_{E_i,Glc}\left[\left(1+S_{Glc}/K_{E_i,S}\right)^{n_i}+L_i/\left((1+A_{Glc}/K_{E_i,A}\right)^{n_i}\right]}{E_{i,Glc}\,S_{Glc}/K_{E_i,S}\left(1+S_{Glc}/K_{E_i,S}\right)^{n_i-1}} \\ f_2 &= \frac{J_{E_i,Act}\left[\left(1+S_{Act}/K_{E_i,S}\right)^{n_i}+L_i/\left((1+A_{Act}/K_{E_i,A}\right)^{n_i}\right]}{E_{i,Act}\,S_{Act}/K_{E_i,S}\left(1+S_{Act}/K_{E_i,S}\right)^{n_i-1}} \quad , \end{aligned} \tag{S14}$$

with substrate $S = PEP$ and activator $A = FBP$ for $i = 3$, and substrate $S = FBP$ and activator $A = PEP$ for $i = 4$. Then, $K_{E_i,S}$ is given implicitly by

$$K_{E_i,S}: \quad 0 = f_1 - f_2 \quad , \tag{S15}$$

and v_{E_i} is given by

$$v_{E_i} = f_1 \quad . \tag{S16}$$

These equations reduce the four–dimensional parameter spaces to two–dimensional solution spaces.

Finally, the kinetic equation for the enzyme E_2 (subproblem 6) contains four uncertain parameters, with two measurements to constrain them. We chose to express $v_{E_2,f}$ and $v_{E_2,r}$ as functions of $K_{E_2,PEP}$ and $K_{E_2,FBP}$, such that

$$\begin{aligned} v_{E_2,f} = &\left[\frac{E_{2,Act}\;PEP_{Act}}{E_{2,Glc}\;PEP_{Glc}} J_{E_2,Glc}\left(1+\frac{FBP_{Glc}}{K_{E_2,FBP}}+\frac{PEP_{Glc}}{K_{E_2,PEP}}\right)+\right. \\ &\left.+J_{E_2,Act}\left(1+\frac{FBP_{Act}}{K_{E_2,FBP}}+\frac{PEP_{Act}}{K_{E_2,PEP}}\right)\right]\left[\frac{E_{2,Act}}{K_{E_2,FBP}}\left(\frac{PEP_{Act}\;FBP_{Glc}}{PEP_{Glc}}-FBP_{Act}\right)\right]^{-1} \\ v_{E_2,r} = &\frac{K_{E_2,PEP}}{PEP_{Glc}}\left[v_{E_2,f}\;\frac{FBP_{Glc}}{K_{E_2,FBP}}-\frac{J_{E_2,Glc}}{E_{2,Glc}}\left(1+\frac{FBP_{Glc}}{K_{E_2,FBP}}+\frac{PEP_{Glc}}{K_{E_2,PEP}}\right)\right] \quad . \end{aligned} \tag{S17}$$

These equations reduce the four–dimensional parameter space to a two–dimensional solution space.

Statistical analysis of the solution space

In this section, we discuss the statistical analysis we used to identify the parameters that most critically determine the system response. To identify these parameters, we take advantage of the two response families introduced in the main paper and compare the parameter combinations that belong to response family A with those that belong to response family B. If a parameter does not critically determine the system response, its value should be equally distributed in both response families. To test whether a parameter is equally distributed in both response families, we compare the distribution of the sampled parameter values between the two families using the two–tailed student's t-test with equal variances. This test results in one p–value per parameter. The p–value is the probability, under the null hypothesis that both parameter sets A and B are drawn from the same distribution, of observing a value as extreme or more extreme of the test statistic

$$t = \frac{\bar{a}\ \bar{b}}{\sqrt{\frac{\sigma}{n_a} + \frac{\sigma}{n_b}}} \quad , \tag{S18}$$

with $\bar{a}$ and $\bar{b}$ the means of the sampled parameter values in both response families, σ the pooled standard deviation, and n_a and n_b the sample sizes. The lower a parameter's p–value is, the more statistically different is the distribution of the sampled parameter values between the two response families A and B.

Table S2 lists the p–values of all parameters. Most parameters exhibit a high p–value and are therefore not suspected to critically shape the system response. Only few parameters exhibit low p–values, and among these, the free parameter $K_{E_2,FBP}$ and the dependent parameter $v_{E_2,f}$ exhibit extremely low p–values. Therefore, the values of these two parameters critically determine to which response family a parameter combination belongs.

In addition, Table S2 lists the mean and standard deviations for each parameter. Most parameters, such as L_{E_i}, exhibit large standard deviations, which implies that these parameter values are only poorly determined. However, some parameters, such as v_{E_3}, v_{E_4} and k_{cat,E_3} exhibit very narrow standard deviations, which implies that the values of these parameters are fairly well determined. The distributions of the parameters with low p–values are significantly different between the two response families.

Table S2: Statistical data (mean values, standard deviations, and p–values) of the distributions of the sampled parameter values in response families A and B.

Free parameter	mean±std family A	mean±std family B	p–value
n_{Cra}	1.37±0.84	1.43±0.96	0.55
$K_{Cra,FBP}$	2.87±0.75	3.00±0.76	0.17
L_{E_1}	$5{\cdot}10^5{\pm}2{\cdot}10^6$	$8{\cdot}10^5{\pm}2{\cdot}10^6$	0.28
$K_{E_1,PEP}$	5.09±2.90	5.18±3.18	0.82
$K_{E_1,Acetate}$	5.14±2.84	4.77±2.70	0.30
L_{E_5}	$6{\cdot}10^5{\pm}2{\cdot}10^6$	$6{\cdot}10^5{\pm}2{\cdot}10^6$	0.88
$K_{E_5,PEP}$	7.07±1.68	6.13±1.49	$1{\cdot}10^{-5}$
$K_{E_5,Glucose}$	5.11±2.90	5.18±2.84	0.86
L_{E_3}	$7{\cdot}10^4{\pm}2{\cdot}10^5$	$8{\cdot}10^4{\pm}2{\cdot}10^5$	0.72
$K_{E_3,PEP}$	5.09±2.79	4.73±2.72	0.32
L_{E_4}	$2{\cdot}10^6{\pm}2{\cdot}10^6$	$2{\cdot}10^6{\pm}2{\cdot}10^6$	0.52
$K_{E_4,FBP}$	0.61±0.26	0.69±0.25	0.01
$K_{E_2,PEP}$	4.91±2.85	5.62±2.65	0.05
$K_{E_2,FBP}$	5.48±2.59	0.85±0.56	$3{\cdot}10^{-145}$
Dependent parameter	mean±std family A	mean±std family B	p–value
v_{E_1}	52.5±133	53.2±126	0.96
v_{E_3}	5.43±0.50	5.42±0.51	0.93
v_{E_4}	$1.16{\pm}7{\cdot}10^{-3}$	$1.16{\pm}8{\cdot}10^{-3}$	0.88
v_{E_5}	$1.17{\pm}9{\cdot}10^{-3}$	$1.17{\pm}9{\cdot}10^{-3}$	0.88
$K_{E_1,Cra}$	2.17±8.18	2.23±7.67	0.96
$K_{E_3,Cra}$	0.09±0.10	0.09±0.10	0.95
$K_{E_4,Cra}$	3.25±0.80	3.12±1.01	0.22
$K_{E_5,Cra}$	2.79±0.69	2.68±0.87	0.22
k_{cat,E_1}	$2{\cdot}10^4{\pm}2{\cdot}10^5$	$1{\cdot}10^4{\pm}6{\cdot}10^4$	0.95
k_{cat,E_5}	$3{\cdot}10^5{\pm}1{\cdot}10^5$	$2{\cdot}10^5{\pm}7{\cdot}10^5$	0.70
k_{cat,E_3}	$0.06{\pm}9{\cdot}10^{-4}$	$0.06{\pm}9{\cdot}10^{-4}$	0.65
$K_{E_3,PEP}$	0.11±0.10	0.11±0.10	0.65
k_{cat,E_4}	184±677	254±626	0.41
$K_{E_4,PEP}$	0.30±0.15	0.32±0.14	0.20
$v_{E_2,f}$	7.59±1.85	4.48±0.36	$3{\cdot}10^{-151}$
$v_{E_2,r}$	5.86±4.61	24.2±22.1	$3{\cdot}10^{-9}$

Further graphical illustration of the results of the statistical analysis

Figure S1 shows Fig. 3c of the main text viewed along the $K_{E_2,PEP}$–axis. From this perspective, it can be seen that a combination of roughly $K_{E_2,FBP} < 3$ and $v_{E_2,f} < 6$ is necessary for the emergence of an attractive second steady state (gray dots).

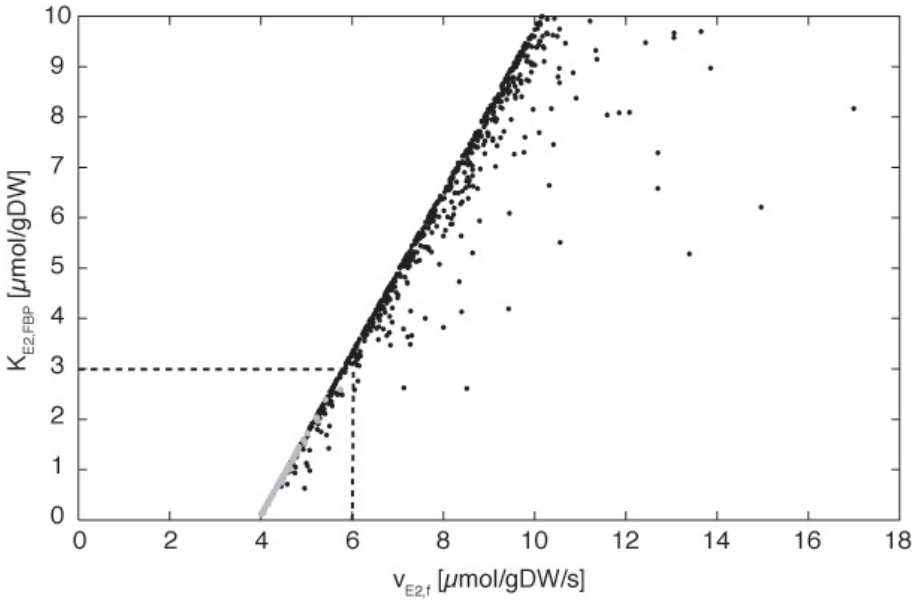

Figure S1: This figure shows the sampled parameters shown in Fig. 3c of the main text viewed along the $K_{E_2,PEP}$–axis. Black dots denote parameter combinations that belong to response family A; gray dots denote those that belong to response family B. The emergence of an attractive second steady state, i.e. a response belonging to family B, is only possible if the parameters $K_{E_2,FBP}$ and $v_{E_2,f}$ lie within a region that is roughly delimited with dashed lines.

References

1. Endy,D., Kong,D., Yin,J. (1996) Intracellular Kinetics of a Growing Virus: A Genetically Structured Simulation for Bacteriophage T7, *Biotech Bioeng.*, **55**, 375-389.

2. Fuhrer,T., Fischer,E., Sauer,U. (2005) Experimental Identication and Quantication of Glucose Metabolism in Seven Bacterial Species, *J Bacteriol.*, **187**, 1581-1590.

3. Lowry,O.H., Carter,J., Ward,J.B., Glaser,L. (1971) The Effect of Carbon and Nitrogen Sources on the Level of Metabolic Intermediates in *Escherichia coli*, *JBC*, **246**, 6511-6521.

4. Oh,M.-K., Rohlin,L., Kao,K.C., Liao,J.C. (2002) Global Expression Profiling of Acetate-grown *Escherichia coli*, *JBC*, **277**, 13175-13183.

5. Zhao,J., Baba,T., Mori,H., Shimizu,K. (2004) Effect of *zwf* gene knockout on the metabolism of *Escherichia coli* grown on glucose or acetate, *Metab Eng.*, **6**, 164-174.